



King's Research Portal

DOI:

[10.1016/j.jclinepi.2015.02.011](https://doi.org/10.1016/j.jclinepi.2015.02.011)

Document Version

Early version, also known as pre-print

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Phillips, R., Gandhi, M., Bun Cheung, Y., Findlay, M. P., Maung Win, K., Hoa Hai, H., Mo Yang, J., Rey Lobo, R., Chee Soo, K., & K.H. Chow, P. (2015). Summary scores captured changes in subjects' QoL as measured by the multiple scales of the EORTC QLQ-C30. *Journal of Clinical Epidemiology*, 68(8), 895-902.
<https://doi.org/10.1016/j.jclinepi.2015.02.011>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Summary scores capture changes in subjects' QoL as measured by the multiple scales of EORTC-QLQ-C30

Rachel **Phillips**^{a, b, *}, Mihir **Gandhi**^{a, b, c}, Yin Bun **Cheung**^a, Michael P **Findlay**^d, Khin Maung **Win**^e,
Hoang Hoa **Hai**^f, Jinmo **Yang**^g, Rolley Rey **Lobo**^h, Khee Chee **Soo**^{b, i}, Pierce KH **Chow**^{b, i, j},

^a Biostatistics Department, Singapore Clinical Research Institute, 31 Biopolis Way, Nanos #0201, 138669, Singapore

^b Office of Clinical Sciences, Academia, 20 College Road, Singapore Duke-NUS Graduate Medical School, Singapore

^c Department of International Health, School of Medicine, University of Tampere, Tampere, Finland

^d Faculty of Medical and Health Sciences, The University of Auckland, Private Bag 92019, Auckland Mail Centre, Auckland 1142, New Zealand

^e Department of Hepatology, Yangon GI and Liver Centre, 91-193, 30th Street (Upper Block), Pabedan Township, 11141 Yangon, Myanmar

^f Choray Hospital, 201B Nguyen Chi Thanh Street, District 5, Ho Chi Minh City, Vietnam

^g Department of Internal Medicine, Catholic University Medical College, 93-6 Chi-dong, Paldal-ku, Suwon, 442-720, Korea

^h Department of Liver and Pancreatic-Biliary Surgery, Davao Doctors Hospital, Quirino Street, Davao City, The Philippines

ⁱ Department of Surgical Oncology, National Cancer Center Singapore, 11 Hospital Drive, 169610, Singapore

^j Dept. Hepatopancreatobiliary and Transplant Surgery, Singapore General Hospital, Outram Road, 169608, Singapore

Email address of authors:

RP: rachel.phillips@scri.edu.sg

MG: mihir.gandhi@scri.edu.sg

YBC: yinbun.cheung@scri.edu.sg

MPF: mp.findlay@auckland.ac.nz

KMW: dr.khinmaungwin@gmail.com

HHH: hai_hh@cityinternationalhospital.com

JMY: jmyangdr@catholic.ac.kr

RRL: rrlobo8@gmail.com

KCS: admskc@nccs.com.sg

PC: pierce.chow.k.h@nccs.com.sg

***Corresponding author:**

Rachel Phillips

Singapore Clinical Research Institute

31 Biopolis Way, Nanos #02-01, Singapore 138669.

Telephone number: +65 6508 6754, Fax number: +65 6508 8317

Email address: rachel.phillips@scri.edu.sg

ABSTRACT

Objective: To examine the performance of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30) global health status/quality of life (QoL) scale and two summary scores to detect changes in the QoL profile over time, according to changes in the individual scales.

Study Design and Setting: Data came from 167 clinical trial patients with unresectable (advanced) hepatocellular carcinoma. The global health status/QoL scale of the questionnaire contained two items: overall health and overall QoL. Nordin and Hinz proposed summary scores for the questionnaire. A mixed effect model was fitted to estimate trends in scores over time.

Results: Predominantly the individual scale scores declined over time, however the global health status/QoL score was stable (rate of change=-0.3 per month, 95% CI: -1.2, 0.6). Nordin's summary score, which gave equal weight to the 15 questionnaire scales and Hinz's summary score, which gave equal weight to the 30 questionnaire items, showed a statistically significant decline over time, 3.4 (95% CI: -4.5, -2.4) and 4.2 (95% CI: -5.3, -3.0) points per month, respectively.

Conclusion: In contrast to the global health status/QoL scale the summary scores proposed by Nordin and Hinz detected changes in subjects' QoL profile described by the EORTC QLQ-C30 individual scales.

Keywords: EORTC QLQ-C30; cancer; advanced hepatocellular carcinoma; response shift; quality of life; scoring procedure

Running title: Summary measures for the EORTC QLQ-C30

Word count: 4560

ClinicalTrials.gov number: NCT00041275

What is new?

Key findings

- Cross-sectionally, the global health status/QoL scale did not have strong correlation with the functioning, symptom and finance scales of the EORTC QLQ-C30.
- Longitudinally, the global health status/quality of life scale disagreed with the majority of the functioning and symptom scales of the EORTC QLQ-C30.
- Nordin's and Hinz's summary scales summarised the changes in overall QoL profile according to changes in functioning and symptom scales in terms of showing a clinically and statistically significant decline.
- The summary scores based on the sum of individual items/scales had better agreement, both cross-sectionally and longitudinally, with the QoL profile captured by the EORTC QLQ-C30 individual scales than the global health status/quality of life scale.

What this adds to what is known?

- Our findings support the concerns raised about use of the global QoL scale as the overall QoL measure; and support the theory that the global QoL scale is more subject to response shift, whereas the scales based on specific items are less subject to these influences.

What is the implication and what should change now?

- Non-specific, global items to measure overall QoL profile captured by the EORTC QLQ-C30, especially longitudinally should be used with caution.
- Nordin's summary score and Hinz's summary score have potential to serve as an overall QoL measure that summarises the profile captured by the EORTC QLQ-C30 individual scales.
- QoL rated by a global rating continue to remain relevant when one wishes to capture patients' perceptions about their overall level of QoL.

1. INTRODUCTION

The preservation of quality of life (QoL) is an important goal in the management of cancer, particularly when the treatment is given with palliative intent or has a high level of toxicity [1, 2]. QoL instruments measure patients' own perceptions of their well-being in various areas such as physical, psychological, social, financial, and somatic [3, 4]. An important characteristic of a QoL instrument is that it can detect changes in well-being of multiple dimensions and hence overall QoL [5]. However, assessing change in overall QoL using a multi-dimensional QoL instrument is challenging as individual dimensions may change in different directions. A summary measure indicating aggregate status of a multi-dimensional QoL profile is desired to evaluate the change in overall QoL over time. This is particularly important in clinical trials, which are often expected to pre-specify a primary end-point.

The EORTC QLQ-C30 is a widely-used instrument for measuring QoL in patients with cancer [3]. The results of a recent systematic review found it to be reliable and recommended it for use in both clinical trials and clinical practice [6]. The QLQ-C30 includes 30 items for 15 dimensions/scales: 5 functional scales (physical, role, cognitive, social and emotional functioning), 3 symptom scales (fatigue, nausea/vomiting and pain), 5 single-item symptom scales (dyspnoea, sleep disturbances, appetite loss, constipation, and diarrhoea), a single-item scale for the perceived financial impact of the disease and treatment, and a global health status/QoL scale comprised of two items (global QoL scale for brevity). The two items of the global QoL scale rate a patient's overall health and overall QoL [7]. The EORTC does not recommend any overall summary of all the item/scale scores. Instead, it suggests using the global QoL scale as the overall summary measure for QoL [7].

Global ratings for QoL are useful in allowing patients to express their own perceptions of the concept; however they can be too simplistic [8]. Concerns about using the 2-item based global QoL scale as the overall QoL measure have been raised by various researchers [9-11]. There are two conceptual issues. Firstly, the use of only two items and ignoring the other 28 items in the QLQ-C30 to characterize overall QoL may lead to loss of precision [12]. Secondly, it is possible that the global QoL scale indicates a better outcome (or worse outcome) whilst the majority of the individual scales indicate worse outcomes (or better outcomes). This would lead to difficulty in interpretation and making conclusions. In particular, Hinz et al. hypothesized that the global QoL scale may be more

affected by the 'response shift' phenomenon than the specific scales, possibly leading to seemingly different trajectories over time [11]. Concerns have also been raised regarding the statistical efficiency of the QLQ-C30 global scale to detect change over time [14]. Hence, there have been suggestions, and counter-arguments, about fully utilizing all 30 items to generate a summary score for overall QoL [9-11, 13].

Nordin et al. suggested a summary score based on the sum of all scales of the QLQ-C30, except the scale for financial problems [9]. Hinz et al. suggested a summary score based on the sum of all items of the QLQ-C30 [11]. Gundy et al. proposed a two-factor model, one for physical and one for mental health, as a basis for deriving summary scores [15]. However, Gundy et al. have not yet proposed an algorithm for deriving summary scores. Nevertheless, no methods for deriving an overall summary score have been endorsed by the EORTC nor is there a consensus among researchers.

Hepatocellular carcinoma (HCC) is expected to affect physical functioning, role functioning, symptoms of fatigue, pain, dyspnea, insomnia, and appetite loss [16]. Patients with unresectable HCC have limited life expectancy, which is associated with poor functional status and distress due to symptoms [17, 18]. Drawing on data from a longitudinal multi-centre clinical study in unresectable (advanced) HCC patients, this paper aimed to examine the performance of the global QoL scale, Nordin's summary score and Hinz's summary score in relation to the individual scales of the QLQ-C30 and in relation to time. We examined whether (i) the baseline score of the global QoL scale and the summary scores of Nordin's and Hinz's were at least moderately correlated with the baseline scores of the functional, symptom and finance scales of the QLQ-C30, (ii) the rates of change in the global QoL scale and the two summary scores were at least moderately correlated with the rates of change in the functional, symptom and finance scales in a longitudinal setting, and (iii) the global QoL scale and the two summary scores declined over time in a degree similar to the majority of the functional, symptom and finance scales.

2. METHODS

2.1. Design and assessments

Data was drawn from a multi-centre randomized controlled double-blind study that recruited patients between February 2002 and November 2006. The study aim was to compare efficacy and safety of megestrol acetate with placebo in patients with unresectable HCC. The primary trial end-point was overall survival and the key secondary end-points were safety and QoL. The trial analysis found no clinically meaningful effect of study treatment on overall survival or the global QoL scale as measured by the QLQ-C30 [17]. Hence this analysis combined the treatment and placebo groups.

Patients were recruited from 7 centres from 6 Asia-Pacific nations: Singapore, Myanmar, New Zealand, Philippines, South Korea and Vietnam. Details of the trial protocol and findings have been reported previously [17]. Patients were randomized (allocation ratio of 1:2) to either placebo or megestrol acetate 320 mg/day for 12 months, in addition to best supportive care. Baseline and monthly assessments included a QoL assessment using the QLQ-C30. In Singapore, where multilingualism is common, participants chose to use the English or Chinese version of the questionnaire. Participants in Vietnam, South Korea and Myanmar used the Vietnamese, Korean and Burmese versions, respectively. In New Zealand and the Philippines, the English versions were used. Before the clinical examination of the visits, patients were asked to fill in the QLQ-C30 questionnaire if they were literate. Otherwise a research coordinator administered the interview. The option of interviewer administration was allowed as per the QLQ-C30 administration guidelines [19]. The protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki and was approved by the Institutional Review Boards of the participating institutions. Written informed consent was obtained before enrolment.

2.2. Scoring algorithms

The QLQ-C30 functional, symptom and finance scale items are scored on a 4-point scale. These 4 levels include: “Not at all”, “A little”, “Quite a bit” and “Very much”. For the present study symptom items and scales were recoded so that a higher score indicated reduced symptomology. Therefore, all items and scales were coded so that a higher score indicated lower outcomes i.e. higher scores indicated a healthier level of functioning or better QoL or lower symptomology/problems. The raw scores (mean of item-level scores) of the scales were transformed to scores that ranged from 0-100 [9]. In the case of missing items if at least half of the items from a scale had been answered then the missing value was replaced with the mean of the items that were

completed, otherwise the scale score was set to missing. For single item measures, the scale score was set to missing if the item was not answered [7]. Patients that had a score for each of the QLQ-C30 scales/items at baseline were included in the analyses.

The global QoL scale: The global QoL scale is comprised of 2 items which are scored from 1 to 7 with a score of 1 representing “very poor” and 7 indicating “excellent”. Similar to the functional and symptom scales, the mean of two item-level scores was transformed to a score that ranged from 0-100 and a higher score indicated a better QoL.

Nordin’s summary score: Nordin et al. suggested a summary score based on the sum of scales except the scale regarding financial problems [9]. However we deemed the financial scale to be of value in our setting so we included it. The summary score was calculated by summing the transformed scores of all the 15 scales (including the global QoL scale), and then further transformed to the 0-100 scale. Nordin et al. proposed that missing values were handled as per the EORTC manual [7]. Therefore in the case of missing items, the value was replaced with the mean of the remaining items of the corresponding scale when at least 50% of the scale was answered. If any of the items remained missing and hence the corresponding scale was missing, the summary score was set to missing.

Hinz’s summary score: Hinz et al. suggested a summary score as the mean value of all 30 items [11]. It was calculated by first transforming all the item-level scores to the 0-100 range. The summary score was then calculated as the mean of the transformed item-level scores (including the items on the global QoL scale). In the case of missing items, the value was replaced with the mean of the remaining items of the corresponding QLQ-C30 scale when at least 50% of the scale was answered [6]. If any of the items remained missing the summary score was set to missing.

2.3. *Statistical analysis*

The relationship between the global QoL scale and the score of each of the functional, symptom and finance scales at baseline was examined using the Spearman’s correlation coefficient (ρ). The baseline global QoL scale was expected to at least moderately correlate with the baseline

score of each functional/symptom/finance scale. Correlations of between 0.4 and 0.59 were deemed moderate and correlations higher than or equal to 0.6 were deemed strong [20].

The number of repeated QoL assessments per patient varied substantially due to poor survival in this patient group (median survival 2 months; 12-month survival rate 3%). The mixed effect model is robust and fully utilizes all available data including those who have only one measurement, therefore it is suitable for the assessment of trajectories in the presence of incomplete observations [21, 22]. A mixed effects model for the global QoL scale and each of the functional/symptom/finance scales was fitted to estimate the trends over time. It was expected that the global QoL scale and the specific scales would show a trend to decrease over time in this study population. The model was of the form:

$$y_{ij} = (b_0 + \beta_{0i}) + (b_1 + \beta_{1i}) \text{ month}_{ij} + e_{ij}$$

where b_0 and b_1 indicate the group's mean level of QoL at baseline and rate of change in mean QoL per month, β_{0i} and β_{1i} indicate the i^{th} patient's deviation from the intercept and slope, respectively, as compared to the group average. They are known as random intercepts and random slopes in mixed models. β_{0i} and β_{1i} were assumed to be normally distributed with mean 0 and variance to be estimated and were allowed to be correlated.

The individual random intercept and slope (rate of change) for each patient were calculated using the Best Linear Unbiased Predictor (BLUP) [21, 22]. The relationship between the individual estimates of the slope for the global QoL scale and individual estimates of the slope for the specific scales were examined using the Spearman's correlation coefficient (ρ). It was expected that the individual rate of change in the global QoL scale would be at least moderately correlated with the individual changes in each of the specific scales.

Nordin's summary score and Hinz's summary score were analyzed similar to the global QoL scale. In addition we ran a sensitivity analysis to examine the extent the correlation coefficients between the specific scales/items and the summary scores were inflated due to the scales and items already being included in the summary scores. We examined this possibility by calculating the correlation

coefficients between scales/items and summary scores excluding any common items from the summary scores.

3. RESULTS

3.1 Descriptive summary

A total of 167 patients (90.3% of the trial participants) completed the QLQ-C30 at baseline and were included in the analysis. Median number of QLQ-C30 questionnaires completed per patient was 3 (25th percentile 2 and 75th percentile 4).

Table 1 shows the demographic and clinical characteristics of the patients at baseline. Median age was 56.1 years (range 20.1 to 80.9 years). The majority of patients were male (86.2%) and most were recruited from centres in Myanmar (29.9%), Singapore (22.2%) and Vietnam (22.2%). The majority of the participants were experiencing symptoms that affected daily functioning: 85.6% were of grade 1 or above on the Eastern Cooperative Oncology Group (ECOG) performance status.

Table 2 shows the summary scores of the EORTC global QoL, summary scores and functional/symptom/finance scales at baseline. Mean scores across both summary scores and scales were towards the upper end of the scale indicating a higher level of QoL and a higher level of functioning and lower level of symptomology. The global QoL score sat in the centre of the scale.

3.2 Correlation with functional/symptom/finance scales

Table 3 shows the correlation coefficients between the global QoL scale, Nordin's summary score, Hinz's summary score and each of the specific scales of QLQ-C30 at baseline. The global QoL scale showed moderate, positive correlation with physical functioning, role functioning, fatigue, pain and appetite loss symptom scales (a higher score meant better outcome for all scales; each $p \geq 0.4$, each $p\text{-value} < 0.05$). Nordin's summary score showed moderate-to-strong, positive correlation with all of the functional scales and symptom scales (each $p \geq 0.4$, each $p\text{-value} < 0.05$), except weak correlation with constipation and diarrhoea symptom scales. Similar results were observed for the correlation between Hinz's summary score and functional/symptom scales at baseline as just described for Nordin's summary score.

Table 4 shows the correlation coefficients between the individual random slope (rate of change over time) for the global QoL scale, Nordin's summary score, Hinz's summary score and the corresponding individual random slope for each of the specific scales, obtained from the mixed effects model. Over time the global QoL scale positively and moderately correlated with the physical functioning, emotional functioning and fatigue symptom scales (each $\rho \geq 0.4$, $p\text{-value} < 0.01$), and negatively and moderately correlated with the diarrhoea symptom scale ($\rho = -0.4$, $p\text{-value} < 0.01$). The rate of change of Nordin's summary score showed moderate-to-strong, positive correlation with physical, emotional, cognitive functioning scales and fatigue, dyspnoea, insomnia symptom scales (each $\rho \geq 0.4$, $p\text{-value} < 0.01$). There was moderate-to-strong, negative correlation with the pain, appetite loss and diarrhoea symptom scales (each $\rho \leq -0.4$, $p\text{-value} < 0.01$). Similar results were observed for correlation between the rate of change in Hinz's summary score and that for the specific scales as just described for Nordin's summary score.

Calculating the correlation coefficients between scales/items and summary scores, excluding any common items from the summary scores, gave systematically lower correlations compared to the main findings but not dramatically so (supplementary tables 1 and 2).

3.3 *Trajectory over time*

The physical, role and emotional functioning scales and the fatigue, dyspnoea and insomnia symptom scales all showed a statistically significant decline over time (Figure 1). In particular, the mean physical functioning scale score declined by 7.7 points per month (95% CI: -9.6 to -5.8). Conversely the appetite loss and diarrhoea symptom scales both showed a trend to increase over time. Despite statistical non-significance, four of the remaining six scales showed a decline over time. Despite the decline in most of the specific scale scores, the global QoL score was stable over time ($b_1 = -0.3$; 95% CI: -1.2 to 0.6). In contrast, Nordin's and Hinz's summary score showed a statistically significant decline over time, by 3.4 points (95% CI: -4.5 to -2.4) and 4.2 points (95% CI: -5.3 to -3.0) per month respectively.

4. DISCUSSION

We examined the performance of three different overall QoL measures, namely the global QoL scale, Nordin's summary score and Hinz's summary score, for QLQ-C30 in unresectable treatment-naïve advanced HCC patients from Asia-Pacific countries.

Our findings showed that the global QoL scale performed in an unexpected way longitudinally and, to a lesser extent, cross-sectionally. Cross-sectionally, at baseline, the global QoL scale had positive, moderate correlation with only physical and role functioning and fatigue, pain and appetite symptom scales. In contrast, not surprisingly, the Nordin's and Hinz's summary scores showed strong correlation with all the functioning scales and moderate-to-strong correlation with most of the symptom scales at baseline. Sensitivity analyses suggested the robustness of these findings. Longitudinally, although most of the functioning and symptom scales, including two of the only five scales that showed moderate correlation with the global QoL scale at baseline, were declining during the study period, the global QoL scale was stable over time (rate of change=-0.3 per month; p-value=0.48). In contrast, the Nordin's and Hinz's summary scales did summarize the overall profile of change in terms of showing a clinically and statistically significant decline.

A stable trajectory for the global QoL scale during the follow-up, even though the majority of functional and symptom scales were showing declines in QoL domains, may have several potential explanations. Firstly, it is theoretically possible that global QoL covered more domains than the present set of functioning and symptom scales did, and these unobserved domains tended to improve over time in this patient group. However, we find it difficult to accept this explanation, given the broad coverage of the QLQ-C30 although the QLQ-C30 was not specifically developed for patients with terminal diseases. Secondly, there might be a subset of patients who see their overall QoL and/or health as always good, perhaps reflecting dispositional optimism [23]. Or some patients might have adapted to their disease. That is, patients' internal standard (e.g., priorities and expectations) changed and they may view their QoL differently over time, a phenomenon referred to as 'response shift' [24]. In the present context, as patients approached the final stage of disease, they might have reached the 'acceptance' stage of psychological response [25]. The two items of the global QoL scale of the QLQ-C30 were more abstract than the specific scales in the questionnaire [11]. The findings are consistent with the hypothesis made by Hinz et al. that the global QoL scale is more subject to response shift

(and/or dispositional optimism), whereas the scales based on more concrete and specific items were less subject to these influences [11].

The use of non-specific, global items to measure quality of life is common in clinical research. Some of them are part of a more comprehensive questionnaire, e.g. the global QoL items in the QLQ-C30. Some of them are standalone measures, e.g. the QoL Uniscale [26]. The present findings put forwards a cautionary note to the use of such items when one wishes to summarise the overall QoL profile captured by the EORTC QLQ-C30, especially in longitudinal settings.

Although the majority of QoL domains declined in this patient group, the pain symptom score was stable over time and the appetite loss symptom score showed statistically significant improvement during the follow-up. Severe pain is not a common feature of advanced HCC (unless there are bone metastases, which are uncommon). The terminal stages are characterized by liver failure: ascites, jaundice and encephalopathy (comatose). Another possible explanation is that the patients might have been on opioid analgesic for managing their severe pain before and during the study [16]. Pharmacologically, randomized controlled trials have demonstrated that megestrol acetate significantly improves appetite [17, 27]. Nearly two-thirds of the patients in our study were treated with megestrol acetate, which may be a reason for improved appetite loss during the study. The use of an overall summary (or a global QoL scale) alone may miss such specific insights. Nevertheless, we consider this a reason to use specific scales in addition to, not instead of, an overall summary measure. This is especially important in clinical trials that require pre-specification of a primary end-point and when the intervention is hypothesized to have benefits to multiple aspects of QoL.

Our study showed that both Nordin's summary score and Hinz's summary score have the potential to serve as an overall QoL measure that summarizes the profile captured by the QLQ-C30 individual scales. Both of them had moderate to strong correlation in the expected direction with the majority of functional and symptom scales associated with HCC [16]. Both of them were consistent with the majority of the functioning and symptom scales in showing a downward QoL trajectory. Despite similar findings about the performance of Nordin's and Hinz's summary scores, there is a formal difference between them. Nordin's summary score gives equal weight to each functional/symptom scale, whereas Hinz's summary score gives equal weight to each item of the QLQ-C30. That is, Hinz's summary score assigns weights to the scales in proportion to the number of

items in the scale. Thus, the physical functioning scale (5 items) has the highest weight, followed by emotional functioning scale (4 items), and so on. As such, the Hinz's scale showed a larger decline over time in QoL in this study than the Nordin's scale. Neither of the two weighting methods is perfect in all situations. Ideally, each patient's weights should be decided by the patient depending on his/her preference to different aspects of QoL and hence they can vary from patient to patient. However, deciding the preference based weights is a complex procedure. Various methods have been discussed in the literature with their merits and demerits and the difficulties involved [28]. In contrast, Nordin's summary score and Hinz's summary score are simple, sufficiently reliable and valid and likely to be useful for many practical purposes.

A limitation of this study was that it had a high 'drop-out' rate and therefore the data was dominated by short-term rather than long-term changes. The high drop-out rate was due to unresectable HCC patients having median life expectancy of less than 3 months [17, 29]. We used a mixed model approach, to assess the rate of change in the QLQ-C-30 scales over time, which is robust and accommodates missing data.

Another limitation was that the patients were from a clinical trial with strict inclusion criteria, which limits the ability to generalize the results to all patients with advanced HCC. Additionally, the questionnaire we used was not developed specifically for terminal disease patients. Further studies of patients with terminal disease should ideally include measurements of QoL concerns specific to this disease stage. The present findings also need replications in patients with non-terminal diseases.

Last but not least, the two item global scale and a summary score of the remaining items are not conceptually comparable [11]. However, the empirical comparison and subsequent discrepancy identified are of scientific value.

In a research setting with no pre-specified interest in a specific QoL component, an overall QoL measure is helpful. Firstly, overall QoL scores help avoid multiple comparisons and maintain the type 1 error without impacting the sample size. Secondly, an overall score is more useful in facilitating clinical decision making than multiple scales that may be in conflict with each other. For example, regarding the use or non-use of an intervention.

For practical purposes, when an overall score is suitable one can use these findings to choose between the global score and the summary scores, dependent on the specific aim of the question. In our opinion, when one wishes to obtain a single summary that reflects the overall QoL profile captured by the EORTC QLQ-C30 individual scales, the summary scores are more appropriate. The summary scores are also recommended when one wishes to detect changes within a group or between two groups [14]. However when one wishes to capture the individual patients' perceptions about their QoL then the global score is useful [8].

Longitudinally, the rate of change in the global QoL scale and Nordin's and Hinz's summary measures showed moderate level of negative correlation with some symptom scales, including diarrhoea, pain and appetite loss. We do not understand why this is so. We observed that these symptom scales either showed an improvement over time or no change over time during the study period. It appears that, in this study population, these symptom scales have some unexpected and unexplained characteristics. This deserves further investigation. From a methodological point of view, this study also highlights issues about the examination of validity and responsiveness to change. In this study, the rate of change (i.e. the slope of individual trajectory) of the global QoL scale had positive correlation with the rates of change of the physical functioning (0.52) and fatigue (0.54) scales. It might be tempting, but incorrect, to think that this implies the measures would show similar trends over time. While the physical functioning and fatigue scale scores declined by about 8 and 3 points per month, the global QoL scale was almost constant over time. The positive correlations only indicate that, within the sample, those who had below (above) average of change in physical functioning and fatigue scores also had below (above) average of change in global QoL scores. The positive correlations do not guarantee that, at the level of the sample as a whole group, the global QoL score would change according to changes in the specific scales.

In conclusion, cross-sectionally, the global QoL scale of QLQ-C30 did not have strong correlation with the functioning and symptom scales in patients with unresectable hepatocellular carcinoma. Longitudinally, it disagreed with the majority of the functioning and symptom scales in showing changes over time. The overall summary scores proposed by Nordin et al. and Hinz et al. were better at detecting changes in the profile of QoL captured by the QLQ-C30 individual scales than

the global QoL scale. The choice between the global scale and the summary scores ultimately comes down to the question of interest.

ACKNOWLEDGMENT

The authors thank the team members involved in designing the trial protocol (D Machin, Y Chen, C-H Lim and S-B Tan), Prof. Ngyuen Ba Duc and the National Cancer Institute, Vietnam as a contributing study site and Singapore Clinical Research Institute for providing support for site monitoring and data management of the trial. Financial support for this study was provided by Pivotal Clinical Trials Grant from SingHealth.

REFERENCES

- [1] Bottomley A. The cancer patient and quality of life. *Oncologist* 2002;7(2):120-125.
- [2] Kaasa S, Loge JH. Quality of life in palliative care: principles and practice. *Palliat Med* 2003;17:11–20.
- [3] Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Filiberti A, Flechtner H, Fleishman SB, de Haes JC, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst.* 1993;85(5):365-76.
- [4] Coons SJ, Rao S, Keininger DL, Hays RD. A comparative review of generic quality-of-life instruments. *Pharmacoeconomics.* 2000;17(1):13-35.
- [5] Fayers PM, Machin D. *Quality of Life: The Assessment, Analysis, and Interpretation of Patient-Reported Outcomes.* 2nd edition. Chichester, UK: John Wiley & Sons; 2007.
- [6] Smith AB, Cocks K, Taylor M and Parry D. Most domains of the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire C30 are reliable. *J Clin Epidemiol* 2014; 67: 952-957.
- [7] Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A, on behalf of the EORTC Quality of Life Group. *The EORTC QLQ-C30 Scoring Manual (3rd Edition).* Published by: European Organisation for Research and Treatment of Cancer, Brussels 2001.
- [8] Wright J.G and Feinstein A.R. A Comparative Contrast of Clinimetric and Psychometric Methods for Constructing Indexes and Rating Scales. *J Clin Epidemiol* 1992; 45(11): 1201-1218.
- [9] Nordin K, Steel J, Hoffman K, Glimelius B. Alternative methods of interpreting quality of life data in advanced gastrointestinal cancer patients. *Br J Cancer.* 2001;85(9):1265-72.
- [10] Cheung YB, Goh C, Thumboo J, Khoo KS, Wee J. Variability and sample size requirements of quality-of-life measures: A randomized study of three major questionnaires. *J Clin Oncol* 2005; 23: 4936-4944.
- [11] Hinz A, Einenkel J, Briest S, Stolzenburg JU, Papsdorf K, Singer S. Is it useful to calculate sum scores of the quality of life questionnaire EORTC QLQ-C30? *Eur J Cancer Care (Engl).* 2012;21(5):677-83.

- [12] Lumley T, Simes RJ, Gebski V, Hudson HM. Combining components of quality of life to increase precision and evaluate trade-offs. *Stat Med* 2001; 20(21):3231-49.
- [13] Coens C, Bottomley A, Efficace F, Flechtner H, Aaronson N. Correspondence: variability and sample size requirements for health-related quality-of-life measures: understanding the challenges facing investigators. *J Clin Oncol* 2005;8541-8542.
- [14] King M.T, Bell M.L, Costa D, Butow P and Oh B. The Quality of Life Questionnaire Core 30 (QLQ-C30) and Functional Assessment of Cancer-General (FACT-G) differ in responsiveness, relative efficiency, and therefore required sample size. *J Clin Epidemiol* 2014; 16(1) 100-107.
- [15] Gundy CM, Fayers PM, Groenvold M, Petersen MA, Scott NW, Sprangers MA, Velikova G, Aaronson NK. Comparing higher order models for the EORTC QLQ-C30. *Qual Life Res.* 2012;21(9):1607-17.
- [16] Sun VC, Sarna L. Symptom management in hepatocellular carcinoma. *Clin J Oncol Nurs.* 2008;12(5):759-66.
- [17] Chow PKH, Machin D, Chen Y, Zhang X, Win KM, Hoang HH, Nguyen BD, Jin MY, Lobo R, Findlay M, Lim CH, Tan SB, Gandhi M, Soo KC. Randomised double-blind trial of megestrol acetate vs placebo in treatment-naïve advanced hepatocellular carcinoma. *Br J Cancer* 2011;105: 945-952.
- [18] Yeo W, Mo FK, Koh J, Chan AT, Leung T, Hui P, Chan L, Tang A, Lee JJ, Mok TS, Lai PB, Johnson PJ, Zee B. Quality of life is predictive of survival in patients with unresectable hepatocellular carcinoma. *Ann Oncol.* 2006;17(7):1083-9.
- [19] The EORTC Groups Websites - FAQs. <http://groups.eortc.be/qol/faq?page=4> (accessed on June 5, 2014).
- [20] Campbell MJ, Swinscow TDV. *Statistics at Square One* (9th Edition). BMJ Books, London 2001.
- [21] Fairclough DL. *Design and Analysis of Quality of Life Studies in Clinical Trials*. 2nd Edition. Chapman and Hall/CRC Interdisciplinary Statistics; 2002.
- [22] Cheung YB. *Statistical analysis of Human Growth and Development*. Chapman & Hall/CRC Biostatistics Series; 2014.

- [23] Sprangers MA. Disregarding clinical trial-based patient-reported outcomes is unwarranted: Five advances to substantiate the scientific stringency of quality-of-life measurement. *Acta Oncol.* 2010;49(2):155-63.
- [24] Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: A theoretical model. *Soc Sci Med* 1999; 48(11):1507–1515.
- [25] Kubler-Ross E. *On death and dying*. New York: Scribner; 1969.
- [26] Bowling A. Just one question: If one question works, why ask several? *J Epidemiol Community Health* 2005; 59:342-345.
- [27] Lopez A, Roque Figuls M, Cuchi G. Systematic review of megestrol acetate in the treatment of anorexiacachexia syndrome. *J Pain Symptom Manage* 2004;27:360–369.
- [28] Drummond MF, O'Brien B, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press, Oxford, 1997.
- [29] Pawarode A, Voravud N, Sriuranpong V et al. Natural history of untreated primary hepatocellular carcinoma: a retrospective study of 157 patients. *Am J Clin Oncol* 1998; 21: 386–391.

TABLES

Table 1: Demographic and baseline disease characteristics

Table 2: Descriptive summaries for the global QoL, summary scores and functional/symptom/finance scales of EORTC QLQ-C30

Table 3: Spearman's correlations between baseline global QoL scale and summary scores and functional/symptom/finance scales of QLQ-C30

Table 4: Spearman's correlation between individual rates of change for the global QoL and summary scores and the individual rates of change for function/symptom/finance scales of QLQ-C30

FIGURES

Figure 1: Rate of change per month (random slope from mixed effects models) for each QLQ-C30 functional, symptom and finance scale, the global QoL scale and summary measures